

PATENT

**CLASSIFICATION OF PATIENTS HAVING DIFFUSE LARGE B-CELL
LYMPHOMA BASED UPON GENE EXPRESSION**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No.

60/510,822, filed on October 14, 2003, which is hereby incorporated in its entirety by reference.

GOVERNMENT INTERESTS

[0002] This work was supported at least in part with funds from the federal government under U.S.P.H.S. Grants CA33399 and CA34233, awarded by the National Institutes of Health. The U.S. Government may have certain rights in the invention.

FIELD

[0003] This application relates generally to gene expression in cancerous tissues and, more particularly, to gene expression in diffuse large B-cell lymphoma (DLBCL) tissues and to methods for classifying patients with DLBCL based upon gene expression in DLBCL tissues.

BACKGROUND

[0004] Although combination chemotherapy for the treatment of DLBCL patients has been available for several years, currently, over one-half of all patients do not achieve a durable remission (Vose, *supra*, 1998). Risk stratification of patients has been attempted to identify patients in which more aggressive treatment may be required. One risk stratification approach has involved use of the International Prognostic Index (IPI), which is based upon 5 clinical criteria (The International Non-Hodgkin's Lymphoma Prognostic Factors Project, *N. Engl. J. Med.* 32:987-993, 1993). However, the IPI has not provided an accurate prediction of survival in a substantial number of patients.

SUMMARY

[0005] Accordingly, the present inventors have succeeded in developing an approach for stratifying DLBCL patients at the molecular level based upon gene expression in DLBCL tissues. The approach involves correlating expression values of a plurality of genes in tumor samples from patients having DLBCL to classification characteristics of the disease, such as, for example, overall patient survival. A set of genes can be selected from the plurality of genes based upon the expression of the selected genes showing a correlation to the classification characteristics. The relationship developed from this correlation can then allow patient classification by measuring expression of the selected genes in a tumor sample from a patient and comparing with expression values obtained in the correlation study. The approach can be applied not only to DLBCL, but also to other cancers as well as non-cancerous diseases.

[0006] Thus, in various embodiments, the present invention can involve methods for classifying a patient or patients having DLBCL into groups based upon classification characteristics. The methods can comprise measuring expression of a plurality of genes, in a tumor sample from a patient and correlating tumor expression values to normalized reference expression values obtained for the plurality of genes from DLBCL patients stratified in the classification groups. In various aspects of this embodiment, the method can predict patient survival based upon the selected plurality of genes being predictive of survival by virtue of being identified in DLBCL patients stratified in groups of known overall survival. In various aspects of this embodiment as well as embodiments described below, classification characteristics other than or in addition to overall survival can be used such as, for example, likelihood of successful treatment for various treatments which can be used to select a specific therapy approach for a given patient. Gene expression can be measured by any method that quantifies gene expression

such as real time RT-PCR. Quantification can be relative or absolute quantification or a combination of both as applied to the normalization process, which is discussed more fully below. Briefly, relative quantification references expression of a target gene to a control value for expression such as, for example, expression obtained from a control sample or pretreatment sample or expression of a reference gene. Absolute quantification is based upon an internal or external calibration curve (see for example, Pfaff et al., *Nucleic Acid Research* 30:e36, 2002; Livak et al., *Methods* 25:402-408, 2001).

[0007] In various other embodiments, the present invention can involve a method for obtaining a formula for classifying patients having a disease, such as, for example, DLBCL. The method comprises correlating normalized expression values of a plurality of genes in tumor samples obtained from patients having the disease to at least one known classification characteristic of the disease. In various aspects of this embodiment, the method can predict patient survival and the classification characteristic of the disease can be overall survival. Gene expression can be measured by any method which quantitates gene expression such as real time RT-PCR. The plurality of genes can be at least two, at least three, at least four, at least five or all of the genes *LMO2*, *BCL-6*, *FN1*, *CCND2*, *SCYA3* and *BCL-2*. Additional genes can also be included.

[0008] The present invention, in various embodiments, can also involve kits for classifying a patient having DLBCL into classification groups, such as, for example, groups predictive of the probability of survival of the patient. The kits contain assays for measuring expression of a plurality of genes in a tumor sample from a patient having DLBCL. The normalized expression of the plurality of genes in tumor samples from DLBCL patients stratifies the patients into classification groups. The assays in the kits can comprise real time RT-PCR

assays. The kits can also contain software for using the expression data so as to simplify the assignment of patients to classification groups.

[0009] In various embodiments, the present invention can also involve a method for predicting survival in a patient having DLBCL. The method comprises measuring in a sample containing tumor cells from the patient, expression of a plurality of genes and determining whether normalized expression of the genes indicates increased or decreased probability of survival. The plurality of genes can be at least three, at least four, at least five or all of the genes *LMO2*, *BCL-6*, *FN1*, *CCND2*, *SCYA3* and *BCL-2*. Additional genes can also be included. In one aspect, determining can involve determining whether normalized expression of the three or more genes matches expression criteria indicative of increased probability of survival, compared to expression in reference cells. The reference cells can be non-cancerous cells from the patient or cells other than DLBCL tumor cells obtained from sources other than the patient such as, for example, Raji cells. The expression criteria can be selected from the group consisting of increased expression of *LMO2*, increased expression of *BCL-6*, increased expression of *FN1*, decreased expression of *CCND2*, decreased expression of *SCYA3* and decreased expression of *BCL-2*. In various aspects of this embodiment, the reference cells can be Raji cells. Gene expression can be measured by any of a number of methods such as, for example, cDNA or cRNA microarray test, tissue microarray test or real time RT-PCR.

[0010] In various of the embodiments above, normalized expression can comprise values calculated by one or both of calculating the ratio of expression values of the target gene and an endogenous reference gene and calculating the ratio of expression values of the target gene to expression of the same gene reference cells with or without normalization to the endogenous reference gene. The endogenous reference gene can be a housekeeping gene such as, for

example, *PGK1* or *GAPDH*. The reference cell line can be a Raji cell line. Reference stratification of patients based upon expression values can be generated using univariate Cox proportional hazards analysis with classification, such as, for example, overall survival as dependent variable. Moreover, the methods can use IPI scores in addition to the gene expression information obtained.

[0011] In various of the embodiments above, gene expression in a patient can be compared to gene expression in reference DLBCL patients of known survival using the formula:

$$Z = (A \times LMO2) + (B \times BCL6) + (C \times FN1) + (D \times CCND2) + (E \times SCYA3) + (F \times BCL2)$$

[0012] The terms LMO2, BCL6, FN1, CCND2, SCYA3 and BCL2 can be log base 2 of normalized expression values for genes *LMO2*, *BCL-6*, *FN1*, *CCND2*, *SCYA3* and *BCL-2*, respectively. In various embodiments A can be about -0.03, B can be about -0.2, C can be about -0.2, D can be about 0.03, E can be about 0.2 and F can be about 0.6. Using these values, a Z value of less than about -0.06 can indicate high probability of survival, a Z value of from about -0.06 to about 0.09 can indicate medium probability of survival and a Z value of greater than about 0.09 can indicate low probability of survival. In various aspects of this embodiment, A can be about -0.0273, B is about -0.2103, C can be about -0.1878, D can be about 0.0346, E can be about 0.1888 and F is can be about 0.5527. Using these values, a Z value of less than about -0.063 indicates high probability of survival, a Z value of from about -0.063 to about 0.093 indicates medium probability of survival and a Z value of greater than about 0.093 indicates low probability of survival.

[0013] Application of the methods of the present invention to clinical practice allows identification of patients who are unlikely to be cured by conventional therapy and in whom investigational approaches would be justified in an effort to improve their outcome.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Figure 1 illustrates univariate analysis of expression of 36 genes using overall survival as a dependent variable, wherein the genes are ranked based on their predictive power (univariate score) with negative score associated with longer overall survival while positive univariate score associated with shorter overall survival, the dashed lines (at a univariate score of 1.5) representing a significance threshold of $p < 0.05$.

[0015] Figure 2 illustrates the development of the 6 gene model showing (A) Kaplan-Meier estimates of overall survival in the 66 DLBCL cases analyzed by quantitative RT-PCR with TaqMan[®] probe-based assays in which dotted lines represent 95% confidence intervals and (B) Kaplan-Meier curves of overall survival in the tertiles (low, medium and high) defined by a prediction model based on the weighted expression of 6-genes (*LMO2*, *BCL-6*, *FN1*, *CCND2*, *SCYA3* and *BCL-2*) in which the significance measures are based on log-likelihood estimates of the p-value, treating the model as a continuous variable or as a class (first and second p-values, respectively).

[0016] Figure 3 illustrates the external validation of the performance of the 6-gene model on data from (A) oligonucleotide microarrays showing in the Left panel, Kaplan –Meier estimates of overall survival for the 58 DLBCL cases reported by Shipp et al. (Shipp et al., *supra*, 2002) in which dotted lines represent 95% confidence intervals and in the right panel, Kaplan-Meier estimates of overall survival of 58 patients when subdivided into tertiles (low, medium and high) using the 6-gene prediction model, the significance measures being based on log-likelihood estimates of the p-value treating the model as a continuous variable or as a class (first and second p-values, respectively) and (B) cDNA microarrays showing a similar analysis of data from the 240 DLBCL cases reported by Rosenwald et al. (Rosenwald et al., *supra*, 2002).

[0017] Figure 4 illustrates the improvement the 6-gene model adds to the International Prognostic Index showing Kaplan–Meier estimates of overall survival for each IPI group (scores 0-1, 2-3, 4-5) of patients reported by Rosenwald et al. (Rosenwald et al., *supra*, 2002) when subdivided into tertiles (low, medium and high) using the 6-gene prediction model in which the significance measures are based on log-likelihood estimates of the p-value treating the model as a continuous variable or as a class (first and second p-values, respectively)(n = 11, 39 and 32 for top, middle and bottom tertiles, respectively, of low IPI score plots; n = 8, 48 and 52 for top, middle and bottom tertiles, respectively, of medium IPI score plots and n = 2, 16 and 14 for top middle and bottom tertiles, respectively, of high IPI score plots).

DETAILED DESCRIPTION

[0018] The present invention, in various embodiments, can involve methods for classifying patients having a disease into groups based upon gene expression values from a plurality of genes. The disease can be DLBCL or other cancers or a non-cancerous disease.

[0019] Classification groups or stratification groups for patients having DLBCL can involve any of a variety of features of the disease, in particular, various aspects that characterize the severity of the disease into groups based upon morbidity or mortality of the patients having the disease. One measure of mortality is “overall survival” sometimes referred to as “survival rate”. The term “overall survival” refers to the percentage of subjects in a study who have survived for a defined period of time, usually measured from the time of diagnosis although it can also be measured from the time of initiation of treatment. Overall survival time of DLBCL patients as referenced herein, is calculated from the date of the diagnosis until death or last follow-up examination.

[0020] Inasmuch as DLBCL patients normally receive various treatments for the disease, overall survival time can mean survival time following chemotherapy. Chemotherapy can be Anthracycline-based chemotherapy and such anthracycline-based chemotherapy, as used herein, is intended to refer to the use of at least one anthracycline-class compound in chemotherapy treatment. As a non-limiting example, doxorubicin is an anthracycline-class compound used for treating non-Hodgkin's lymphoma and this compound can be used in a combination treatment of cyclophosphamide, doxorubicin, vincristine and prednisone (Vose, *supra*, 1998).

[0021] In various embodiments, the disease DLBCL can be identified in patients prior to applying the methods of the present invention. Methods of diagnosing DLBCL are well known in the art such as, for example, the use of histologic and immunologic criteria (see for example, Harris et al, *Blood* 84:1361-1392, 1994; The Non-Hodgkin's Lymphoma classification Project, *Blood* 89:3909-3918, 1997). After identification, the methods of the present invention can be used to classify patients having the disease.

[0022] In various embodiments, the methods of the present invention can also be used in determining whether DLBCL is present in a patient and in distinguishing of DLBCL from other diseases as well as in monitoring of the disease status or the recurrence of the disease, and in determining a preferred therapeutic regimen for the patient. Gene expression in DLBCL tumors can thus, be used in the diagnosis of DLBCL patients. Assessing the gene expression profile of DLBCL tumors can, in certain instances, provide a diagnostic basis for identifying disease aggressiveness and tumor progression (Lossos et al, *Int. J. Hematol.* 77:321-329, 2003). Thus, in various embodiments, classification of patients into survival probability groups can constitute the classification of patients into subsets of DLBCL diseases having different clinical prognoses.

[0023] Identification of patterns of gene expression can form the basis for understanding tumorigenesis at the molecular level as well as the underlying mechanisms that may contribute to disease aggressiveness and tumor progression (Lossos et al., *Int. J. Hematol.* 77:L321-329, 2003). Thus, evaluation of gene expression related to DLBCL can provide a more meaningful approach to understanding the disease than has been available in histologic or other clinical tests that have attempted to classify patients with DLBCL. Gene expression involves transcription of genomic DNA to form RNA's and ultimately proteins in the cell. Assessing gene expression can be done by determining cellular RNA or protein levels in a cell. Numerous methods for measuring gene expression at the RNA or protein level are known. Non-limiting examples of methods that measure RNA include Northern blotting, nuclease protection assays, DNA microarrays, serial analysis of gene expression, quantitative reverse transcription-polymerase chain reaction (RT-PCR), differential-display RT-PCR, massively parallel signature sequencing and the like. In particular, measurement of gene expression at the RNA level can be performed using real-time quantitative RT-PCR assay such as exonuclease-based assays, for example, TaqMan® assays. Non-limiting examples of methods of measuring protein expression levels include mass spectrometry, two-dimensional gel electrophoresis, antibody microarrays, tissue microarrays, ELISA, radioimmunoassay, immuno-PCR and the like.

[0024] In various embodiments, the methods of the present invention can be used to identify the pattern of gene expression in DLBCL and to determine the relationship to various aspects of DLBCL such as, for example, disease prognosis. A number of genes have been suggested to be related to DLBCL (see for example Alizedeh et al., *Nature* 403:503-511, 2000; Shipp et al., *supra*, 2002; Rosenwald, et al., *supra*, 2002 and Table 1 below). These and other genes can be evaluated using various methods of the present invention to assess the relationship

of gene expression to disease prognosis such as overall survival in a population of individuals having DLBCL and to determine the prognosis of an individual having the disease. In particular, *BCL-6* has been shown to predict survival in DLBCL patients using real-time RT-PCR methods (Lossos et al., *Blood* 98:945-951, 2001). Thus, in various embodiments, *BCL-6* can be one of the genes used to classify DLBCL patients in overall survival groups.

[0025] In various embodiments gene expression values can be normalized to provide more accurate quantification and to correct for experimental variations. In various aspects of the invention, the calculation of gene expression values from the real-time RT-PCR tests can involve generating C_t (threshold cycle) values for target gene and an endogenous reference gene RNAs from control and experimental samples; determining nanogram amounts of each RNA using calibration standard curves; calculating the ratio of target and endogenous gene reference RNA; and calculating the ratio of nanograms target gene RNA in control and experiment samples. The endogenous reference RNA can be that of a housekeeping gene (see for example, Lossos et al, *Leukemia* 17:789-795, 2003). In particular, phosphoglycerate kinase 1 (*PGKI*) or glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) can be used as the endogenous reference RNA. Calibration standard curves can be generated using cDNA from Raji cells or from Universal Human Reference RNA (Stratogene, La Jolla, CA). Raji cells can also be used for determining control target gene RNA and endogenous gene RNA. Normalization aspects of the calculations can comprise one or both of calculating the ratio of expression values of the target gene and an endogenous reference gene and calculating the ratio of expression values of the target gene to expression of the same gene in a reference cell line with or without normalization to the endogenous reference gene. Other normalization methods that correct for experimental variation can also be used (for review see Freeman et al, *BioTechniques* 26:112-125, 1999).

[0026] The normalized gene expression values can be transformed to log-base 2 values. Further evaluation can then be performed by comparing the transformed values with selected classification criteria using various statistical methods. In constructing a survival prediction model, the normalized gene expression can be compared to overall survival as estimated using the product-limit method of Kaplan-Meier with comparisons based upon the log-rank test. Cox proportional hazards analysis with overall survival as the dependent variable can then be performed. Genes with an absolute univariate Cox score between -1.5 and 1.5 can then be analyzed by multivariate regressions analysis using a Cox proportional hazards regression model with overall survival as the dependent variable.

[0027] The invention can be further understood by reference to the examples which follow.

EXAMPLE 1

[0028] This example illustrates the selection of genes potentially predictive of overall survival and the performance of quantitative RT-PCR on the selected genes.

[0029] Thirty-six genes were selected for inclusion in the study as shown in Table 1.

Table 1: Sources of supporting evidence for panel of 36 prognostic genes assessed in this study

Genes (total=36 total) [†]	Reference
<i>ICAM1/CD54</i>	Terol et al., <i>Ann Oncol</i> 14:467-74, 2003.
<i>PAX5</i>	Krenacs et al., <i>Blood</i> 92:1308-16, 1998
<i>Ki-67</i>	Miller et al., <i>Blood</i> 83:1460-6, 1994
<i>CD44</i>	Drillenburger et al., <i>Leukemia</i> 13:1448-55, 1999

Genes (total=36 total) [†]	Reference
<i>P53</i>	Ichikawa et al., <i>N Engl J Med</i> 337:529-34, 1997; Koduru et al., <i>Blood</i> 90:4078-91, 1997
<i>BCL-2</i>	Gascoyne, et al, <i>Blood</i> 90:244-51, 1997; Kramer et al., <i>J Clin Onco</i> 14:2131-8, 1996; Hermine et al., <i>Blood</i> 87:265-72, 1996; Hill et al., <i>Blood</i> 88:1046-51, 1996
<i>BIRC5/SURVIVIN</i>	Adida et al., <i>Blood</i> 96:1921-5, 2000
<i>BCL-6</i>	Lossos et al., <i>Blood</i> 98:945-951, 2001; Barrans et al., <i>Blood</i> 99:1136-43, 2002
<i>PRDMI</i>	Shaffer et al., <i>Immunity</i> 13:199-212, 2000
<i>HGAL</i>	Lossos et al., <i>Blood</i> 101:433-40, 2003
<i>SCYA3</i>	Shaffer et al., <i>Immunity</i> 13:199-212, 2000
<i>SCYA3</i>	Shaffer et al., <i>Immunity</i> 13:199-212, 2000
<i>CCND1</i>	Shaffer et al., <i>Immunity</i> 13:199-212, 2000
<i>CCND2</i>	Shaffer et al., <i>Immunity</i> 13:199-212, 2000
<i>LMO2, LRMP, CD10, MYBL1/A-MYB, BCL7A, PIK3CG, CR2, CD38, SLAM, WASPIP, CFLAR, SLA, IRF4, PMS1, HGAL, BCL-6, BCL-2</i>	Alizadeh et al., <i>Nature</i> 403:503-11, 2003 ^{††}
<i>NR4A3, PDE4B</i>	Shipp et al. <i>Nat Med</i> 8:68-74, 2002
<i>FN1, PLAUI, HLA-DQA1, HLA-DRA, EEFI1A1L4, NPM3, MYC, BCL-6, HGAL</i>	Rosenwald et al., <i>N Engl J Med</i> 346:1937-47, 2002

[†] Some of the genes are present in more than one source and are thus repeated in the table. We also included three genes that are known targets of *BCL-6* (*PRDMI*, *SCYA3*, *CCND2*) based on work by Shaffer et al, given the prominence of *BCL-6* in DLBCL.

^{††} In addition to representatives from the ~71 genes employed by Alizadeh et al, we also included genes based on a reanalysis of the dataset using SAM .

[0030] The expression of each of these had previously been reported to predict DLBCL survival, either in single gene studies or in the analysis of large data sets derived from microarray studies. In addition, we applied Significance Analysis of Microarrays (Tusher et al., *Proc Natl Acad Sci U S A* 98:5116-21, 2001) – a supervised method for the identification of genes significantly associated with survival – to the dataset of Alizadeh et al. (Alizadeh et al., *supra*, 2000), to detect and recover any significant genes missed in the exploratory analyses employed by the authors.

[0031] Tumor specimens from patients newly diagnosed with DLBCL were obtained during the course of diagnostic procedures at Stanford University medical center between the years of 1975 and 1995. Specimens were stored as previously reported. All the DLBCL tumors had the histological appearance of centroblastic large cell lymphomas demonstrating diffuse pattern of involvement without evidence of residual follicles. All patients were treated with an anthracycline containing chemotherapy regimen and had clinical follow up at Stanford University Hospital. A total of 66 primary DLBCL specimens fulfilled these inclusion criteria. Staging information was obtained for all the patients according to the Ann Arbor system. The IPI score was able to be determined for 59 of these patients.

[0032] For each of these 36 genes and a pair of internal controls for input mRNA (PGK1 and GAPDH), we measured gene expression using quantitative RT-PCR, based on primer and probe sets shown in Table 2. We assayed the expression of each gene in each of the 66 patient specimens relative to that in a reference RNA sample. Isolation of RNA, its quantification and the RT reactions were performed as previously reported (Lossos et al., *Blood* 101:433-40, 2003; Lossos et al., *Leukemia* 17:789-95, 2003).

Table 2: Primer and probe sequences employed for panel of 36 genes surveyed.

Gene	Accession No. (Assay I.D. No.)*	Sequences		SEQ. ID. NO.
<i>ICAM1</i> / <i>CD54</i>	NM_000201 (Hs00277001_m1)	Probe	CTGTTCCAGGACCTG	1
		Forward Primer	ACGCTGAGCTCCTCTGCTACTC	2
		Reverse Primer	CACAGATGTCTGGGCATTGC	3
<i>PMS1</i>	NM_000534 (Hs00153333_m1)	Probe	CATAGTTCTCCAGTTTAAC	4
		Forward Primer	TGGATGCTGTGTGCCACAA	5
		Reverse Primer	CGTTATCTCGCACCTCAATTTATC	6
<i>p53/TP53</i>	NM_000546 (Hs00153340_m1)	Probe	CTGGCTGCCAATCC	7
		Forward Primer	GCTTTCCACGACGGTGACA	8
		Reverse Primer	TGACTGCGGCTCCTCCAT	9
<i>BCL2</i>	NM_000633 (Hs00153350_m1)	Probe	ACAAAGGCATCCCAGCC	10
		Forward Primer	ACCTGCACACCTGGATCCA	11
		Reverse Primer	ACAGCCAGGAGAAATCAAACAGA	12
<i>BIRC5</i> / <i>SURVIVIN</i>	NM_001168 (Hs00153353_m1)	Probe	CCTTTGCAATTTG	13
		Forward Primer	CTGGACAGAGAAAGAGCCCAAGAA	14
		Reverse Primer	GGCACGGCGCACTTCT	15

<i>PRDM1</i>	NM_001198 (Hs00153357_m1)	Probe	TGAATCTCACAAACAC	16
		Forward Primer	CCCGGAGAGCTGACAAATGA	17
		Reverse Primer	TCAGTGCTCGGTTGCTTTAGACT	18
<i>BCL6</i>	NM_001706 (Hs00277037_m1)	Probe	CCTTACCATTTGTGAGAAAGT	19
		Forward Primer	GCGAATCCACACAGGAGAGAA	20
		Reverse Primer	TTGTGACGGAAATGCAGGTTAC	21
<i>CCND1/ PRAD1</i>	NM_001758 (Hs00277039_m1)	Probe	TCCATTTGCAGCAGCT	22
		Forward Primer	CCGAGAAAGCTGTGCATCTACAC	23
		Reverse Primer	AGGTTCCACTTGAGCTTGTTCAC	24
<i>CCND2</i>	NM_001759 (Hs00277041_m1)	Probe	ACAGACCTCCAGCATC	25
		Forward Primer	CCCTACATGCCGAGAAATGGT	26
		Reverse Primer	GACCTCTTCTTCGCACCTTCTGTTC	27
<i>CD38</i>	NM_001775 (Hs00277045_m1)	Probe	CTTCTGCAAACCTGC	28
		Forward Primer	GGAGAAAGGACTGCAGCAACA	29
		Reverse Primer	AGCATCACATGGACCCACATCAC	30
<i>CR2</i>	NM_001877 (Hs00153398_m1)	Probe	AGGCACACACCAGTTT	31
		Forward Primer	CCAGCCAGCTGATCAGAAAGAC	32
		Reverse Primer	TCCGCTGAATTCCAAGCAAT	33

<i>Ki-67</i>	NM_002417 (Hs00267195_m1)	Probe	CACTCTCATCAGGGTCAG	34
		Forward Primer	CGTCGTGTCTCAAGATCTAGCTTCT	35
		Reverse Primer	GGACACACGCCCTTCTTTTCAA	36
<i>IRF4</i>	NM_002460 (Hs00277069_m1)	Probe	CCCAGCAGGTTTCAC	37
		Forward Primer	CTACACCATGACAAACGCCCTTACC	38
		Reverse Primer	GGCTGATCCGGGACGCTAGT	39
<i>MYC</i>	NM_002467 (Hs00153408_m1)	Probe	TCCTCCTCAGAGTCGC	40
		Forward Primer	CCCCTGGTGCTCCATGAG	41
		Reverse Primer	GCCTGCCCTCTTTTCCACAGA	42
<i>PDE4B</i>	NM_002600 (Hs00277080_m1)	Probe	TCGCATTTCAGGTCCTT	43
		Forward Primer	CAGGCGTTCTTCTCTCTAGACAACTA	44
		Reverse Primer	GGTCTGCACAGTGTAACCATGTG	45
<i>PIK3CG</i>	NM_002649 (Hs00277090_m1)	Probe	ATGTCCTGAAATTTC	46
		Forward Primer	GGAAAGAAAGACAAAGCCCACACTT	47
		Reverse Primer	GTGATGACGAAGGGCTAGATAAGC	48
<i>SCYA3/ CCL3</i>	NM_002983 (Hs00234142_m1)	Probe	CTGCATCACTTGCTGC	49
		Forward Primer	ATGGCTCTCTGCAACCAGTTCT	50
		Reverse Primer	GCGGTCGGCGGTGCA	51

<i>SLAM</i>	NM_003037 (Hs00234150_m1)	Probe	ACCTGCTCATAAAGC	52
		Forward Primer	CGCTTTTGCCTGCAGTTGA	53
		Reverse Primer	CGTTCTCCTGGGTCTTGTTTAAA	54
<i>WASPIP</i>	NM_003387 (Hs00277097_m1)	Probe	CTCCAGAAATCATATATCC	55
		Forward Primer	TGAGATCCACGGCCAAACAG	56
		Reverse Primer	CTCCCGGTGGCAACAATG	57
<i>CFLAR</i>	NM_003879 (Hs00153439_m1)	Probe	AGACAGAGCTTCTTCG	58
		Forward Primer	TGCCTGATAATCGATTGCATTG	59
		Reverse Primer	CCCAGGGAAAGTGAAAGGTGTCT	60
<i>LMO2</i>	NM_005574 (Hs00277106_m1)	Probe	AGACTATCTCAGGCTTTT	61
		Forward Primer	CAAAC TGGGCCCGGAAGCT	62
		Reverse Primer	ATGCGCAGAGACCGTCTTG	63
<i>LRMP</i>	NM_006152 (Hs00277107_m1)	Probe	AAAGCTTCTTCAGTTTCC	64
		Forward Primer	GCCAGGCACTTCAGCAGAA	65
		Reverse Primer	GGACAGGGAATCATAGTGAGGAAAT	66
<i>SLA</i>	NM_006748 (Hs00277129_m1)	Probe	CCTGCAGTCTGGACAC	67
		Forward Primer	TCACCTTGCGTGAGAAAGACTGT	68
		Reverse Primer	CCCAAGCGGGTTCTCTGT	69

<i>NR4A3</i>	NM_006981 Hs00175077_m1	Probe	CATGATCACAGAAAGAC	70
		Forward Primer	GCCTGCCTGTCAGCACTGA	71
<i>CD10</i>	NM_007289 Hs00153519_m1	Reverse Primer	GCTCTTCGACTCTCTTTGGTTCTTT	72
		Probe	CCCAATAATCCTGAAATT	73
<i>PAX5</i>	NM_016734 Hs00277134_m1	Forward Primer	CTATAGGCCAGAGTATGCGGTTAAC	74
		Reverse Primer	GCAGTGAAAGGCTTCTGAAAACCTC	75
<i>CD44</i>	AJ251595.1 (NM_000610)	Probe	CCAGTGGACACTATGC	76
		Forward Primer	CAGTCCCAGCTTCCAGTCACA	77
<i>CD44</i>	AJ251595.1 (NM_000610)	Reverse Primer	ACGAGCCGGCCGAATC	78
		Probe	CAGCCATTCTGGAATT	79
<i>CD44</i>	AJ251595.1 (NM_000610)	Forward Primer	GCAAAACACAACCTCTGTGTCCTATAA	80
		Reverse Primer	CCAAGAGGGATGCCCAAGATG	81
<i>HGAL/ GCET2</i>	AF521911 (NM_152785) (Hs00277164_m1)	Probe	CCATCCAGGACAAATGT	82
		Forward Primer	CCCAAAACGAAAATGAAAGAAATGT	83
<i>MYBL1/ A-MYB</i>	X66087.1 (Hs00277146_m1)	Reverse Primer	GGGTATAGCACAGCTCCTCTGAGTA	84
		Probe	CAATTGACTGAAGATTC	85
<i>MYBL1/ A-MYB</i>	X66087.1 (Hs00277146_m1)	Forward Primer	AACCAAAACCCTAACACTTCCAAAG	86
		Reverse Primer	TTCTGTCTTCCCATAAACCACTGTT	87

<i>BCL7A</i>	NM_020993 Hs00277139_m1	Probe	CTTGTTTTTGTCAACAAC	88
		Forward Primer	AATCTACAAATGGGTCCCTGTGA	89
		Reverse Primer	ACACTTCTCGTCCCTTGCCTTT	90
<i>PLAU</i>	NM_002658 (Hs00170182_m1)	Probe	CCTGCCAGGGAGAC	91
		Forward Primer	TCTGAAGTCACCACCACCAAAATGCT	92
		Reverse Primer	CGGCCTTGGAGGGAACA	93
<i>NPM3</i>	NM_006993 (Hs00199625_m1)	Probe	CTCATCGTAACAATCTG	94
		Forward Primer	GGCCCTGTGCGGATCAC	95
		Reverse Primer	CTCTCCTCCTCAGAAACATCATTG	96
<i>HLA-DQA1</i>	NM_002122 (N/A)	Probe	CACTGGGAGCCTGAG	97
		Forward Primer	CCTGGACCAGCCTCTTCTGA	98
		Reverse Primer	CAGTCTCTGTGAGCTCTGACATAGG	99
<i>HLA-DRA</i>	NM_019111 (Hs00219578_m1)	Probe	CATCACCTCCATGTGC	100
		Forward Primer	GCAAAAAGCAATGCAGCAGAAC	101
		Reverse Primer	GGATTGTAAATATTGCCAGCTTTGTAA	102
<i>FN1</i>	NM_002026 (Hs00365058_m1)	Probe	ATCCAAGCTCAAGTGGTCC	103
		Forward Primer	CTATGGCCGTGGCATTGG	104
		Reverse Primer	GTGGGAGTTGGGCTGACT	105

<i>EEF1A1L14</i>	NM_001403 (Hs00265885_m1)	Probe	CCAAGGCATGTTAGCAC	106
		Forward Primer	TTTCTGTTGGAAATGGTGACAA	107
		Reverse Primer	GGGTGACTTTCCCATCCCTTGA	108
		Probe	CCCTGGTGACCCAGGC	109
<i>GAPDH</i>	XM_171437 (4326317E)	Forward Primer	AGCCGAGCCACATCGCT	110
		Reverse Primer	TGGCAACAAATATCCACTTACCAGAGT	111
<i>PGK1</i>	NM_000291 (4326318E)	Probe	AAGGTTAAAGCCCGAGCCA	112
		Forward Primer	GGGAAAAGATGCTTCTGGGAA	113
		Reverse Primer	TTGGAAAGTGAAAGCTCGGAAA	114

* Assay I.D. No. of Applied Biosystems, Foster City, CA.

[0033] Expression of mRNA for 36 tested genes and 2 endogenous control genes was measured in each DLBCL specimen with real time PCR using the Applied Biosystems Assays-on-Demand™ Gene Expression Products on an ABI PRISM® 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) as previously reported (Lossos et al., *Leukemia* 17:789-95, 2003). For each gene, 2-4 assays (TaqMan® probe and primer sets) were tested. The probes contain a 6-carboxy-fluorescein phosphoramidite (FAM™ dye) label at the 5' end and a minor groove binder (MGB) and non-fluorescent quencher (NFQ) at the 3' end, and designed to hybridize across exon junctions. The assays are supplied with primers and probe concentrations of 900 nM and 250 nM, respectively. Real-time assays used in this study had high (near 100%) amplification efficiencies.

[0034] No fluorescent signal was generated by these assays when genomic DNA was used as a substrate, validating the assays as measuring mRNA only. The assays were highly reproducible with inter-run variance of less than 0.16 for all the genes. Phosphoglycerate kinase 1 (*PGKI*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) were used as the endogenous RNA /cDNA quantity controls (P/N 4326318E and P/N 4326317E, respectively Applied Biosystems, Foster City, CA). We chose *PGKI* and *GAPDH* based on an analysis of their relatively constant expression in DLBCL tumors. Since the normalization to *PGKI* and *GAPDH* endogenous control genes lead to similar results and conclusions, we present only the data normalized to *PGKI* expression. For calibration and generation of standard curves we used Raji cDNA and/or cDNA prepared from Universal Human Reference RNA (Stratagene, La Jolla, CA). The latter was used for genes with low abundance in Raji cell line (*CCND1*, *CCND2*, *SLA*, *NR4A3*, *CD44*, *PLAU*, and *FNI*). To control for possible variability between different PCR runs performed on different days, expression of all the analyzed and

endogenous control genes was assessed in Raji cell line before, midway and upon completion of the analysis of all the experimental DLBCL specimens. The variance between these 3 runs for all the genes assessed in the Raji cell line was less than 0.16.

[0035] Calculation of normalized gene expression values was performed as follows.

C_t values measured from tumor samples were converted to quantity of RNA expressed in ng/ μ l, by referencing to the standard curve for the gene. For each gene the ratio of the quantity expressed to the quantity of expression of the reference gene, GAPDH was then calculated. For each gene, the same ratio was determined for calibrator RNA obtained from Raji cells or from the Universal Standard Reference. Finally, the ratio obtained from the tumor sample was divided by the ratio obtained for the calibrator cells.

[0036] Gene expression values for each of the 36 genes and 66 patients is shown in

Table 3.

Table 3. Normalized Gene Expression Values Determined in Sixty-Six Patients Referenced to GAPDH and Raji Cells Unless Otherwise Indicated

Gene	Normalized Expression Values		
	Mean	Variance	Standard Deviation
<i>ICAM1/CD54</i>	2.26	8.16	2.86
<i>PMSI</i>	3.26	6.31	2.51
<i>p53</i>	2.58	3.66	1.91
<i>BCL-2</i>	21.23	840.02	28.98
<i>BIRC5/SURVIVIN</i>	1.31	0.93	0.96
<i>PRDM1</i>	32.44	817.94	28.60
<i>BCL-6</i>	5.62	117.79	10.85
<i>CCND1*</i>	0.78	4.04	2.01
<i>CCND2*</i>	4.18	57.02	7.55
<i>CD38</i>	11.01	85.15	9.23
<i>CR2</i>	2.05	17.25	4.15
<i>Ki-67</i>	1.77	1.01	1.00
<i>IRF4</i>	49.66	4641.85	68.13

Gene	Normalized Expression Values		
	Mean	Variance	Standard Deviation
<i>MYC</i>	2.07	7.69	2.77
<i>PDE4B</i>	36.30	1238.24	35.19
<i>PIK3CG</i>	9.20	56.19	7.50
<i>SCYA3</i>	9.72	158.87	12.60
<i>SLAM</i>	1.01	1.23	1.11
<i>WASPIP</i>	6.95	42.93	6.55
<i>CFLAR</i>	23.53	1800.59	42.43
<i>LMO2</i>	7.34	62.07	7.88
<i>LRMP</i>	3.90	8.12	2.85
<i>SLA*</i>	108.59	23782.23	154.21
<i>NR4A3*</i>	8.41	97.50	9.87
<i>CD10</i>	1.51	3.38	1.84
<i>PAX5</i>	6.46	176.48	13.28
<i>M17</i>	1.87	3.11	1.76
<i>MYBL1/A-MYB</i>	3.72	17.33	4.16
<i>BCL7A</i>	2.44	4.31	2.08
<i>CD44 (139)*</i>	5.13	15.53	3.94
<i>PLAU*</i>	6.51	99.49	9.97
<i>NPM3</i>	1.49	2.28	1.51
<i>HLA-DQA1</i>	2.91	7.19	2.68
<i>EEF1A1L4</i>	1.42	0.70	0.84
<i>HLA-DRA</i>	4.43	11.52	3.39
<i>FNI*</i>	2.46	14.57	3.82

*Referenced to Stratagene Universal Reference RNA.

EXAMPLE 2

[0037] This example illustrates the statistical evaluation for developing a survival predictive model.

[0038] The normalized gene expression values were log-transformed (base 2) similar to what is done with hybridization array data.

[0039] Overall survival time of DLBCL patients was calculated from the date of the diagnosis until death or last follow-up examination. Survival curves were estimated using the product-limit method of Kaplan-Meier and were compared using the log-rank test.

[0040] To determine a small list of genes whose expression segregated DLBCL tumors into subgroups with distinct overall survival, we performed a univariate Cox proportional hazards analysis with the overall survival as the dependent variable. Genes with an absolute univariate Cox score >1.5 or <-1.5 were analyzed by a multivariate regression analysis (with and without IPI components) using a Cox proportional hazards regression model with overall survival as the dependent variable. This same model was used to adjust the effects of gene expressions for IPI. p values < 0.05 were considered to be significant. Backward stepwise analysis was also used, to find the minimal set of genes that were predictive. A p -value cutoff of 0.05 was used for deletion of model terms.

[0041] Results of the univariate analysis are shown in Figure 1. The genes were ranked based upon their predictive power (univariate score) with negative score associated with longer overall survival while positive univariate score associated with shorter overall survival. Six genes with absolute univariate score >1.5 (LMO2, BCL-6, FN1, CCND2, SCYA3 and BCL-2) were selected for further analysis. On multivariate Cox regression analysis with DLBCL overall survival as a dependent variable, none of these genes independently predicted overall survival at a statistically significant level, however on backward stepwise analysis, expression of LMO2 correlated with DLBCL overall survival ($p=.011$). Multivariate Cox regression analysis incorporating all the components of IPI together with the expression of these 6 genes disclosed that only LDH was an independent predictor of DLBCL overall survival ($p=.0038$). However, on backward stepwise analysis,

both LDH and LMO2 expression were independent predictors of DLBCL overall survival ($p=.0035$ and $p=.025$, respectively).

[0042] Since this analysis established an inter-correlation between the expressions of these 6 genes and survival, we constructed a model based on a weighted predictor derived from the relative contributions of each gene in the multivariate analysis. The weighted predictor (z) was calculated for each tumor specimen and the tumors were ranked into 3 tertiles: low, medium and high using the -0.63 and 0.093 as cut points (<-0.063 – low risk, between -0.063 to <0.093 , medium risk and >0.093 – high risk groups). The overall survival of these 3 groups was significantly different ($p=.004$) with 5-year survival of 65%, 49% and 15% for the low, medium and high groups, respectively (mean overall survival [95% confidence interval] of 7.1 {5.4 - not achieved}, 9.0 {1.1 - not achieved} and 4.5 {1.2-4.3} years, respectively, Figure 2). Consequently, patients with tumors expressing high levels of LMO2, BCL-6 and FN1 and low levels of CCND2, SCYA3 and BCL-2, survived longer.

[0043] For construction of the survival prediction model, we derived the weighted predictor (Z) from the multivariate analysis for each of the six genes:

$$Z = (-0.0273 \times \text{LMO2}) + (-0.2103 \times \text{BCL6}) + (-0.1878 \times \text{FN1}) + (0.0346 \times \text{CCND2}) + (0.1888 \times \text{SCYA3}) + (0.5527 \times \text{BCL2}).$$

[0044] Thus for example the negative weight on LMO2 means that higher expression correlates with lower risk (longer survival). The positive weight on CCND2 means that higher expression correlates with higher risk (shorter survival).

EXAMPLE 3

[0045] This example illustrates the validation of the survival prediction model.

[0046] To validate the usefulness of the model derived in Example 2, the model was applied to two independent previously published DLBCL gene expression data sets derived from DNA microarray methodology (Shipp et al., *supra*, 2003; Rosenwald et al., *supra*, 2003). Application of the 6 gene prediction model to data from Shipp et al. (Shipp et al., *supra*, 2003) (Figure 3A) and to that of Rosenwald et al. (Rosenwald et al., *supra*, 2002) (Figure 3b) confirmed its ability to predict survival since it could stratify DLBCL cases into 3 subgroups with statistically significant different overall survival ($P=.03$ and $P=.0004$, respectively). Although in the smaller DLBCL cohort reported by Shipp et al., the overall survival of the group in the medium tertile was similar at the 5 year point to that of their high risk tertile, this medium tertile did have an intermediate risk in the larger cohort of patients analyzed by Rosenwald et al. (Rosenwald et al., *supra*, 2002) (Figure 3B).

[0047] We next analyzed whether this prediction model could add to the prognostic value of the IPI. In our own series of 66 patients there were not enough patients in the lowest risk IPI group to achieve statistical significance. But in our patients within the high clinical risk IPI group, the six gene expression model could further subdivide the patients in respect to survival ($P=.006$) (data not shown). We, therefore, tested the model on the larger DLBCL data set derived from microarray analysis reported by Rosenwald et al. (Rosenwald et al., *supra*, 2002) (Figure 4). We used their same three subdivisions of the patients according to the IPI (low, medium and high risk). Within each of these subgroups we further divided the patients according to the 6 gene expression model. In some of these groups the patients numbers were limited. But in each IPI strata we could identify an especially poor surviving group (Fig. 4 blue lines). By combining the lowest surviving tertiles from the medium and

high risk IPI strata, then we identify 30 % of all patients that receive very little benefit from current therapy.

[0048] The present study defined and validated across the published studies a small set of genes whose expression can predict DLBCL survival and which can be measured by a clinically applicable method. To this end, we evaluated side-by side the prognostic significance of 36 representative genes chosen based on the previous reports suggesting their prognostic potential or from our own analysis of the existing microarray data (Table 1). We have designed a prediction model of overall survival consisting of 6 genes that subdivided DLBCL patients into three prognostic groups in our series of 66 patients and in independent groups of 58 and 240 DLBCL tumors analyzed by Shipp et al. (Shipp et al., *supra*, 2002) and Rosenwald et al. (Rosenwald et al., *supra*, 2002), respectively. The validation of our model did not require any adjustments of the published microarray data or any refinements of our gene list. Moreover, this model could further sub-classify DLBCL patients within IPI strata into longer- and shorter-term survivors. The genes comprising this model are present in each of the previously denoted lymphocyte signatures such as germinal (*LMO2* and *BCL-6*), activated B cell (*BCL-2*, *CCND2*, *SCYA3*) and lymph node signatures (*FN1*) (Alizadeh et al., *supra*, 2000; Rosenwald et al., *supra*, 2002). However, the model is independent of these signatures and several genes associated with these signatures do not carry predictive power in our model.

[0049] *LMO2*, *BCL-6* and *FN1* were the genes whose expression correlated with prolonged survival. *LMO2* was first discovered by its homology with the T cell oncogene *LMO1* (Boehm et al., *Proc Natl Acad Sci U S A* 88:4367-71, 1991). It plays an important role in erythropoiesis and angiogenesis presumably through transcriptional regulation (Warren et

al., *Cell* 1994; 78:45-57, 1994; Yamada et al., *Proc Natl Acad Sci U S A* 97:320-4, 2000).

The *LMO2* locus on chromosome 11p13 is the most frequent site of chromosomal translocation in childhood T-cell acute lymphoblastic leukemia (Boehm et al., *supra*, 1991).

LMO2 is expressed in myeloid and erythroid precursors of hematopoietic system and its expression decreases during differentiation. *LMO2* expression is low in resting peripheral B cells, however it is markedly increased in GC lymphocytes (Alizadeh et al., *supra*, 2000).

LMO2 is not expressed in normal T lymphocytes, however following chromosomal translocation, its ectopic expression in thymocytes contributes to the leukemogenesis (Royer-Pokora et al., *Oncogene* 6:1887-93, 1991). Interestingly, in two recently observed cases of leukemia complicating retrovirus based gene therapy of X-linked severe combined immunodeficiency, the vector inserted itself near the *LMO2* gene (Kaiser, *Science* 299:495, 2003). Neither the functional significance of increased *LMO2* expression in GCB lymphocytes nor its potential role in GCB-derived tumors is known.

[0050] The *BCL-6* gene, identified by virtue of its involvement in chromosomal translocations affecting band 3q27, encodes a POZ/Zinc finger sequence-specific transcriptional repressor (Chang et al., *Proc Natl Acad Sci U S A* 93:6947-52, 1996; Kerckaert et al., *Nat Genet* 1993; 5:66-70, 1993; Seyfert et al., *Oncogene* 1996; 12:2331-42, 1996). The *BCL-6* gene is normally expressed in B and CD4⁺ T cells within the germinal center (GC), and it controls GC formation and T-cell-dependent antigen responses (Cattoretti et al., *Blood* 86:45-53, 1995; Dent et al., *Proc Natl Acad Sci U S A* 95:13823-8, 1998; Ye et al., *Nat Genet* 16:161-70, 1997). It is considered one of the hallmarks of the GC and is expressed in NHL whose origin is from GCB lymphocytes. *BCL-6* expression was previously reported to predict DLBCL outcome (Lossos et al., *Blood* 98:945-951, 2001).

[0051] *FN1* is a component of extracellular matrix in the lymph-node signature. Its expression may reflect the response of the lymph node to the tumor cells. Indeed, some cases of DLBCL demonstrate a sclerotic reaction. This gene, together with *BCL-6*, was included in the survival prediction model constructed by Rosenwald et al. (Rosenwald et al., *supra*, 2002).

[0052] In contrast to these 3 genes, *expression of BCL-2, CCND2, SCYA3* correlated with short survival. All of these 3 genes are included in the ABC-like signature (Alizadeh et al., *supra*, 2000). BCL2 protein expression is down-regulated in normal GCB cells, but is frequently up-regulated in NHL by virtue of t(14;18) translocation (Alizadeh et al., *supra*, 2000; Kramer et al., *Blood* 92:3152-62). Overexpression of the BCL2 protein is known to prevent apoptosis. High BCL2 protein expression has been repeatedly shown to be an independent poor prognostic indicator for DLBCL (Gascoyne et al., *Blood* 90:244-51, 1997; Kramer et al., *J Clin Oncol* 14:2131-8; Hermine et al., *Blood* 87:265-72, 1996; Hill et al., *Blood* 88:1046-51, 1996).

[0053] *CCND2* encodes a protein that belongs to the highly conserved cyclin family, whose members are characterized by a dramatic periodicity in protein abundance through cell cycle. This cyclin forms a complex with CDK4 or CDK6 and regulates their activity thus controlling the cell cycle G1/S transition. Consequently, its expression may be associated with higher proliferation rates of the tumors. *SCYA3* is a CC chemokine that recruits inflammatory cells, including lymphocytes, monocytes, eosinophils and mast cells to sites of inflammation (Proost et al., *Int J Clin Lab Res* 26:211-23, 1996). Its function in B cell lymphomas is unknown, but it is mainly expressed in the ABC-like group of DLBCL tumors and its expression in lymphocytes can be induced by B cell receptor stimulation (Alizadeh et

al., *supra*, 2000). Interestingly, the promoter regions of both *CCND2* and *SCYA3* genes contain high-affinity BCL6 binding sites and the expression of these two genes is repressed by BCL6 (Shaffer et al., *Immunity* 13:199-212, 2000). This observation underscores the complex interrelation between the expression of individual genes singularly implicated in DLBCL prognosis (e.g. *HGAL*) (Lossos et al., *Blood* 101:433-40, 2003), however not contributing to the model based on multivariate analysis.

[0054] All references cited in this specification are hereby incorporated by reference. Any discussion of *references* cited herein is *intended* merely to summarize the assertions made by their authors and no admission is made that any reference or portion thereof constitutes relevant prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

[0055] The description of the invention is merely exemplary in nature and, thus, variations that do not depart from the gist of the invention are intended to be within the scope of the invention. Such variations are not to be regarded as a departure from the spirit and scope of the invention.